



Estimating the Number of Regimes of Non-linear Autoregressive Models.

Joseph Rynkiewicz

► To cite this version:

Joseph Rynkiewicz. Estimating the Number of Regimes of Non-linear Autoregressive Models.. Revista Investigacion Operacional, 2013, 34 (2), pp.117-127. hal-00695539

HAL Id: hal-00695539

<https://hal.science/hal-00695539>

Submitted on 9 May 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating the Number of Regimes of Non-linear Autoregressive Models.

J. Rynkiewicz

SAMM Universite Paris 1
90 Rue de Tolbiac, 75013 Paris, France

May 9, 2012

Abstract

Autoregressive regime-switching models are being widely used in modelling financial and economic time series such as business cycles (Hamilton, 1989; Lam, 1990), exchange rates (Engle and Hamilton, 1990), financial panics (Schwert, 1989) or stock prices (Wong and Li, 2000). When the number of regimes is fixed the statistical inference is relatively straightforward and the asymptotic properties of the estimates may be established (Francq and Roussignol, 1998; Krishnamurthy and Rydén, 1998; Douc R., Moulines E. and Rydén T., 2004). However, the problem of selecting the number of regimes is far less obvious and hasn't been completely answered yet. When the number of regimes is unknown, identifiability problems arise and, for example, the likelihood ratio test statistic is no longer convergent to a χ^2 -distribution. In this paper, we consider models which allow the series to switch between regimes and we propose to study such models without knowing the form of the density of the noise. The problem we address here is how to select the number of components or number of regimes. One possible method to answer this problem is to consider penalized criteria. The consistency of a modified BIC criterion was recently proven in the framework of likelihood criterion for linear switching models (see Oltéanu and Rynkiewicz 2012). We extend these results to mixtures of nonlinear autoregressive models with mean square error criterion and prove the consistency of a penalized estimate for the number of components under some regularity conditions.

keywords

time series, switching regimes, mean square error, asymptotic statistic, models selection, multilayer perceptron

1 The model - definition and regularity conditions

Throughout the paper, we shall consider that the number of lags is known and, for ease of writing, we shall set the number of lags equal to one, the extension to l time-lags being immediate.

Let us consider the real-valued time series Y_t which verifies the following model

$$(1) \quad Y_t = F_{\theta_{X_t}^0}(Y_{t-1}) + \varepsilon_t,$$

where

- X_t is an iid sequence of random variables valued in a finite space $\{1, \dots, p_0\}$ and with probability distribution π^0 ;
- for every $i \in \{1, \dots, p_0\}$, $F_{\theta_i^0}(y) \in \mathcal{F}$ and $\mathcal{F} = \{F_\theta, \theta \in \Theta, \Theta \subset \mathbb{R}^l \text{ compact set}\}$ is the family of possible regression functions. We suppose throughout the rest of the paper that $F_{\theta_i^0}$ are sublinear, that is they are continuous and there exist $(a_i^0, b_i^0) \in \mathbb{R}_+^2$ such that $|F_{\theta_i^0}(y)| \leq a_i^0 |y| + b_i^0, (\forall) y \in \mathbb{R}$;
- for every $i \in \{1, \dots, p_0\}$, $(\varepsilon_t)_{t \in \mathbb{Z}}$ is an independent centered noise, independent of $(Y_t)_{t \in \mathbb{Z}}$.

We need the following hypothesis which implies, according to Yao and Attali (2000), strict stationarity and geometric ergodicity for Y_t :

$$(\mathbf{HS}) \quad \sum_{i=1}^{p_0} \pi_i^0 |a_i^0|^s < 1$$

Let us remark that hypothesis **(HS)** does not request every component to be stationary and that it allows non-stationary “regimes” as long as they do not appear too often.

2 Estimation of the number of regimes

Let us consider an observed sample $\{y_1, \dots, y_n\}$ of the time series Y_t . Then, for every observation y_t , the conditional expectation with respect to the previous y_{t-1} and marginally in X_t is

$$E(Y_t | y_{t-1}) = \sum_{i=1}^{p_0} \pi_i^0 F_{\theta_i^0}(y_{t-1}) := g^0(y_{t-1})$$

As the goal is to estimate p_0 , the number of regimes of the model, let us consider all possible conditional expectation up to a maximal number of regimes P , a fixed positive integer. We shall consider the class of functions

$$\mathcal{G}_P = \bigcup_{p=1}^P \mathcal{G}_p, \mathcal{G}_p = \left\{ g \mid g(y_1) = \sum_{i=1}^p \pi_i F_{\theta_i}(y_1) \right\},$$

where $\pi_i \geq \eta > 0, \sum_{i=1}^p \pi_i = 1$.

For every $g \in \mathcal{G}_P$ we define the number of regimes as

$$p(g) = \min \{p \in \{1, \dots, P\}, g \in \mathcal{G}_p\}$$

and let $p_0 = p(g^0)$ be the true number of regimes.

We can now define the estimate \hat{p} as the argument $p \in \{1, \dots, P\}$ maximizing the penalized criterion

$$(2) \quad T_n(p) = \sup_{g \in \mathcal{G}_p} E_n(g) - a_n(p)$$

where

$$E_n(g) = \frac{1}{2} \sum_{t=2}^n (y_t - g(y_{t-1}))^2$$

and $a_n(p)$ is a penalty term.

Convergence of the penalized mean square estimate

For $\lambda > 0$, let us define the generalized derivative function :

$$\begin{aligned} d_\theta^\lambda(Y_t, Y_{t-1}) &= \frac{\frac{e^{-\lambda(Y_t - g(Y_{t-1}))^2} - e^{-\lambda(Y_t - g^0(Y_{t-1}))^2}}{e^{-\lambda(Y_t - g(Y_{t-1}))^2}}}{\left\| \frac{e^{-\lambda(Y_t - g(Y_{t-1}))^2} - e^{-\lambda(Y_t - g^0(Y_{t-1}))^2}}{e^{-\lambda(Y_t - g(Y_{t-1}))^2}} \right\|_2} \\ &= \frac{e^{-\lambda((Y_t - g(Y_{t-1}))^2 - (Y_t - g^0(Y_{t-1}))^2)} - 1}{\|e^{-\lambda((Y_t - g(Y_{t-1}))^2 - (Y_t - g^0(Y_{t-1}))^2)} - 1\|_2} \end{aligned} \quad (1)$$

and let us define $(d_\theta^\lambda)_-(x, y) = \min\{0, d_\theta^\lambda(x, y)\}$.

Several statistical and probabilistic notions such as mixing processes, bracketing entropy or Donsker classes will be used hereafter. For parcimony purposes we shall not remind them, but the reader may refer to Doukhan (1995) and Van der Vaart (2000) for complete monographs on the subject.

The consistency of \hat{p} is given by the next result, which is an extension of Gassiat (2002):

Theorem 1 : Consider the model (Y_t, Y_{t-1}) defined by (1) and the penalized-likelihood criterion introduced in (2). Let us introduce the next assumptions :

(A1) $a_n(\cdot)$ is an increasing function of p , $a_n(p_1) - a_n(p_2) \rightarrow \infty$ when $n \rightarrow \infty$ for every $p_1 > p_2$ and $\frac{a_n(p)}{n} \rightarrow 0$ when $n \rightarrow \infty$ for every p

(A2) the model (Y_k, X_k) verifies the weak identifiability assumption

$$\sum_{i=1}^p \pi_i F_i(y_1) = \sum_{i=1}^{p_0} \pi_i^0 F_i^0(y_1) \Leftrightarrow \sum_{i=1}^p \pi_i \delta_{\theta_i} = \sum_{i=1}^{p_0} \pi_i^0 \delta_{\theta_i^0}$$

(A3) It exists $\lambda > 0$ so that $\{d_\theta^\lambda, \theta \in \Theta\}$ is a Donsker class.

Then, under hypothesis (A1)-(A3), (HS), $\hat{p} \rightarrow p_0$ in probability.

The assumption (A1) is fairly standard for model selection, in the Gaussian case (A1) will be fulfilled by the BIC criterion. Note that the weak identification assumption (A2) does not allowed to use linear regression because the regression functions have to be linearly independents. The assumption (A3) is more difficult to check. First we note:

$$\begin{aligned} \left(e^{-\lambda((Y_t - g(Y_{t-1}))^2 - (Y_t - g^0(Y_{t-1}))^2)} - 1 \right)^2 &= \\ e^{-2\lambda((Y_t - g(Y_{t-1}))^2 - (Y_t - g^0(Y_{t-1}))^2)} - 2e^{-\lambda((Y_t - g(Y_{t-1}))^2 - (Y_t - g^0(Y_{t-1}))^2)} + 1 \end{aligned}$$

So, d_θ^λ is well defined if $E \left[e^{-2\lambda((Y_t - g(Y_{t-1}))^2 - (Y_t - g^0(Y_{t-1}))^2)} \right] < \infty$, but

$$\begin{aligned} (Y_t - g(Y_{t-1}))^2 - (Y_t - g^0(Y_{t-1}))^2 &= \\ (Y_t - g^0(Y_{t-1}) + g^0(Y_{t-1}) - g(Y_{t-1}))^2 - (Y_t - g^0(Y_{t-1}))^2 &= \\ 2\varepsilon(g^0(Y_{t-1}) - g(Y_{t-1})) + (g^0(Y_{t-1}) - g(Y_{t-1}))^2 \end{aligned}$$

where $\varepsilon = Y_t - g^0(Y_{t-1})$ is the noise of the model. So, if the regression functions are bounded, d_θ^λ is well defined if $\lambda > 0$ exists such that $e^{\lambda|\varepsilon|} < \infty$ i.e. ε admits exponential moments. Finally, using the same techniques of reparameterization as in Liu and Shao, 2003 or Oltéanu and Rynkiewicz, 2012, assumption **(A2)** can be shown to be true for mixture of MLP regression models.

3 Conclusion

We have proven the consistency of BIC-like criteria for estimating the number of components in a mixture of non-linear regression. This result can be shown without knowing the form of the density function of the noise, although the weak identifiability assumption excludes linear regression functions. Finally, a more challenging task may be to get a more precise tuning of penalization term which, according to our result, can be chosen among a wide range of functions.

References

- [1] DOUC R., MOULINES E., RYDÉN T. (2004) Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, *The Annals of Statistics*, **32(5)**, 2254-2304
- [2] DOUKHAN P. (1995) *Mixing: properties and examples*, Springer-Verlag, New York
- [3] DOUKHAN P., MASSART P., RIO E. (1995) Invariance principles for absolutely regular empirical processes, *Ann. Inst. Henri Poincaré (B) Probabilités et Statistiques*, **31(2)**, 393-427
- [4] ENGEL CH., HAMILTON J.D. (1990) Long swings in the dollar: are they in the data and do markets know it?, *American Economic Review*, **80(4)**, 689-713
- [5] FRANCO C., ROUSSIGNOL M. (1998) Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum likelihood estimator, *Statistics*, **32(2)**, 151-173
- [6] GARCIA R. (1998) Asymptotic null distribution of the likelihood ratio test in Markov switching models, *International Economic Review*, **39(3)**, 763-788
- [7] GASSIAT E. (2002) Likelihood ratio inequalities with applications to various mixtures, *Ann. Inst. Henri Poincaré*, **38**, 897-906
- [8] HAMILTON J. D. (1989) A new approach to the economic analysis of non-stationary time series and the business cycle, *Econometrica*, **57(2)**, 357-384

- [9] HAMILTON J. D. (1990) Analysis of time series subject to changes in regime, *Journal of Econometrics*, **64**, 307-333
- [10] HENNA J. (1985) On estimating the number of constituents of a finite mixture of continuous distributions, *Ann. Inst. Statist. Math.*, **37**, 235-240
- [11] IZENMAN A.J., SOMMER C. (1988) Philatelic mixtures and multivariate densities, *Journal of the American Stat. Assoc.*, **83**, 941-953
- [12] KERIBIN C. (2000) Consistent estimation of the order of mixture models, *Sankhya: The Indian Journal of Statistics*, **62**, 49-66
- [13] LAM P.-S. (1990), The Hamilton model with a general autoregressive component: estimation and comparison with other models of economic time series, *Journal of Monetary Economics*, **26**, 409-432
- [14] LIU X., SHAO Y. (2003) Asymptotics for likelihood ratio tests under loss of identifiability, *The Annals of Statistics*, **31(3)**, 807-832
- [15] OLTÉANU M., RYNKIEWICZ J. (2012) Asymptotic properties of autoregressive regime-switching models, *ESAIM P.S.*, **to appear**
- [16] SCHWERT G.W. (1989) Business cycles, financial crises and stock volatility, *Carnegie-Rochester Conference Series on Public Policy*, **31(1)**, 83-125
- [17] VAN DER VAART A.W. (2000) *Asymptotic Statistics*, Cambridge University Press
- [18] WONG C.S., LI W.K. (2000) On a mixture autoregressive model, *J. Royal Statist. Soc, Series B*, **62(1)** 95-115
- [19] YAO J.F., ATTALI J.G. (2000) On stability of nonlinear AR processes with Markov switching, *Advances in Applied Probability*, **32(2)**, 394-407